

Introduction to Artificial Intelligence, Fall & Winter 2022  
College of Computer Science, Zhejiang University  
Problem Set 4: Machine Learning (II)

丁尧相

2022 年 12 月 18 日

**Problem 1.**(Logistic regression) 在 Lecture 11 幻灯片的第 15 页, 我们给出了使用对率线性概率模型时, 二分类 logistic loss 的一种形式:

$$L(\mathbf{x}, y; f) = \log(1 + e^{-y\mathbf{w}^T \mathbf{x}}), \quad y \in \{-1, 1\}. \quad (1)$$

而在第 17 页, 我们则给出了同样使用对率线性概率模型时, 一般  $C$  分类 ( $C \geq 2$ ) 下的 logistic loss:

$$L(x, y; f) = - \sum_{c=1}^C \mathbb{I}[y = y_c] \log p(y_c | \mathbf{x}), \quad p(y_c | \mathbf{x}) = \frac{e^{\mathbf{w}_c^T \mathbf{x}}}{\sum_{i=1}^C e^{\mathbf{w}_i^T \mathbf{x}}}. \quad (2)$$

试论证公式 (1) 中定义的二分类损失, 和公式 (2) 在  $C = 2$  时的损失本质上是等价的。

**Problem 2.**(Back propagation) 请参考 Lecture 11 幻灯片第 48 页, 回答下面的问题:

- 画出函数  $f(x, y, a, b) = (x + y)(a + b)$  所对应的计算图。
- 给定  $x = 2, y = 3, a = 4, b = 5$ , 请根据计算图给出利用反向传播算法计算各变量偏导数值的过程。

**Problem 3.**(Convolution) 设输入图像大小为  $84 \times 84 \times 3$ 。请设计两个不同的具有三个卷积 (可带 pooling) 层的网络结构, 即指定每一层卷积的 kernel size, filter 个数, 有无 padding, stride 大小, 有无 pooling 及 pooling 尺寸, 使得通过这三个卷积层后, 输出尺寸为  $12 \times 12 \times 64$ 。

**Problem 4.**(\*\*\*) (Activation function) 在 Lecture 12 中, 我们讲到 gradient vanishing 是妨碍神经网络层数变深的主要障碍之一, 而使用更合适的激活函数可以有效缓解这一问题, 据此, 请回答下面的问题:

- 请说明为什么采用 Relu:  $f(x) = \max(0, x)$  作为激活函数, 比采用 Sigmoid:  $f(x) = \frac{1}{1+e^{-x}}$  能更好地缓解梯度消失问题。

- 假设在 MLP 网络中使用 Sigmoid 作为激活函数。设该神经元激活前的函数表达式为  $g = -\mathbf{w}^T \text{Sig}(g') + b$ , 其中  $\text{Sig}$  即为 Sigmoid 激活函数,  $\mathbf{w}, b$  分别为权重向量和偏置,  $g'$  为上一层神经元激活前的输出向量。试证明  $\mathbf{w}$  每一维对应的偏导数总是同号的。
- 上一问的现象对神经网络的学习有没有负面影响? 为什么?
- 若改为使用 Relu 作为激活函数, 是否存在上面的问题?